

### R Output from Weight Adjustment Procedure

Lines in RED represents text copied and pasted into R from in98adjustedweights.txt (via a word/text processor).

Lines in BLUE represent output from the R statistical software

Key things to remember when starting this procedure.

1. Make sure that R has the correct directory information for your analysis folder. Usually accomplished by placing a "short-cut" of R in the analysis folder. That folder should also include a sub-folder "original\_data", containing files inbr98.txt and siteevaluation98.txt
2. Load the package "psurvey.analysis" into R
3. Output files will appear in the analysis folder

```
> # File: in98adjustweights.txt
> # Purpose: 1998 survey design for Upper Wabash Basin adjust weights
> #           using Design file, frame information, and Site Evaluation file
>
> # Frame Stream length from RF3. Length includes all Strahler orders and
possibly
> # both perennial and Non-perennial streams
> # The design documentation has 7663.497 km of stream; with 305.347 km in 5th-
7th order
> framelen <- 7358.150
> names(framelen) <- 'Upper Wabash'
>
> # Import Design File for 1998 Upper Wabash Basin
> DesignFile <- read.table('original_data\\inrb98.txt',sep='\t',header=TRUE)
> names(DesignFile)
[1] "SiteID"          "SiteName"        "County"          "RF3RCHID"
[5] "MAP24K"          "MAP100K"         "MAP250K"         "Long.dd"
[9] "Lat.dd"          "Stratum"         "MDCaty"          "Weight"
[13] "StrahlerOrder"   "OECO.96"
> nrow(DesignFile)
[1] 100
> DesignFile[1:10,]
  SiteID      SiteName    County      RF3RCHID    MAP24K
1 INRB98-001    Mud Creek RANDOLPH 5120103 26 3.43 Maxville
2 INRB98-002          HUNTINGTON 5120101 542 0.00 Andrews
3 INRB98-003    EIGHTMILE CR  WELLS 5120101 17 9.63 Ossian
4 INRB98-004    SALAMONIE R  JAY 5120102 147.04 Pennville
5 INRB98-005 HOAGLUND DITCH  WHITE 5120106 29 0.00 Monon
6 INRB98-006 West Honey Creek HOWARD 5120107 145 0.00 Russiaville
7 INRB98-007    Bridge Creek CARROLL 5120105 12 0.64 Delphi
8 INRB98-008          CASS 5120104 160 0.00 Logansport
9 INRB98-009       EEL R  CASS 5120104 2 2.23 Twelve Mile
10 INRB98-010    TIPPECANOE R MARSHALL 5120106 19 5.35 Mentone
  MAP100K  MAP250K  Long.dd  Lat.dd  Stratum  MDCaty  Weight StrahlerOrder
1  MUNCIE  MUNCIE 85.00357 40.21902 3 21 180.57800 1
2  WABASH  MUNCIE 85.60551 40.85739 3 21 180.57800 1
3  WABASH  MUNCIE 85.17342 40.87987 3 22 57.73040 2
4  MUNCIE  MUNCIE 85.14647 40.48114 3 24 26.41177 4
5 LOGANSPORT DANVILLE 86.96295 40.78734 3 23 29.60582 3
6 LA FAYETTE DANVILLE 86.27908 40.43969 3 22 57.73040 2
7 LOGANSPORT DANVILLE 86.63316 40.54491 3 21 180.57800 1
8 LOGANSPORT DANVILLE 86.32255 40.77944 3 21 180.57800 1
9 LOGANSPORT DANVILLE 86.18221 40.81585 3 24 26.41177 4
```

```

10      KNOX CHICAGO 86.09641 41.22948      3      24  26.41177      4
    OECO.96
1      55
2      55
3      55
4      55
5      54
6      55
7      55
8      56
9      55
10     56
> # What are the MDcaty and StrahlerOrder present?
> table(DesignFile$MDCaty)

21 22 23 24
25 25 26 24
> table(DesignFile$StrahlerOrder)

  1  2  3  4
25 25 26 24
>
> # Import 1998 Site Evaluation file from Indiana
> SiteEval <- read.delim('original_data\\siteevaluation98.txt')
> names(SiteEval)
[1] "SiteIDF"           "Basin"            "HUC14Code"
[4] "StreamName"        "StrahlerOrder"      "StatusB"
[7] "Comments"          "BiologicalSampleDate" "QHEI.Score"
[10] "QHEI.Status"       "IBI.Score"         "IBI.Status"
> nrow(SiteEval)
[1] 100
> # Check codes for StatusB - Biological Site Evaluation status
> #           and for StatusWC - Water Chemistry Evaluation status
> table(SiteEval$StatusB)

LD   NT   OT   PB  SCNB   TS   UK
19    9    2    7   14    48    1

> # Note on code meanings
> # TS - target and sampled site
> # NT - site was non-target (dry, non-wadeable,...)
> # LD - landowner denied access to site
> # PB - physical barrier - unable to reach site
> # OT - target site for which a sample was not collected
> # UK - Unknown whether site is target or non-target (shouldn't happen)
> # SCNB - target site Sampled for Chemistry but Not Biology (sampled fewer
biology sites)
>
> # Create designstatus file
> tmp <- match(SiteEval$SiteIDF, DesignFile$SiteID)
> # Check to see if every site in SiteEval is found in DesignFile
> table(is.na(tmp))

FALSE
100
> # if result from above has any TRUE values then sites are missing
> # Don't repeat SiteID variable from SiteEval file

```

```

> designstatus <- data.frame(DesignFile[tmp,],SiteEval[,-4])
> names(designstatus)
 [1] "SiteID"           "SiteName"          "County"
 [4] "RF3RCHID"         "MAP24K"            "MAP100K"
 [7] "MAP250K"          "Long.dd"           "Lat.dd"
[10] "Stratum"          "MDCaty"            "Weight"
[13] "StrahlerOrder"    "OECO.96"           "SiteIDF"
[16] "Basin"             "HUC14Code"         "StrahlerOrder"
[19] "StatusB"          "Comments"          "BiologicalSampleDate"
[22] "QHEI.Score"       "QHEI.Status"        "IBI.Score"
[25] "IBI.Status"
>
> # Add equal area x,y coordinates to be used in variance estimation
> # Note that longitude must be changed to be negative
> tmp <- marinus(designstatus$Lat.dd,-designstatus$Long.dd)
> designstatus$xmarinus <- tmp[, 'x']
> designstatus$ymarinus <- tmp[, 'y']
>
> # Adjust weights to match frame length
> # Determine which sites to include in weight adjustment.
> # All 100 sites were evaluated so all should be used.
> sites <- rep(TRUE,100)
> designstatus$final.wt <- adjwgt.fcn(sites,designstatus$Weight ,
+ designstatus$Basin,framelen)
> # Check out weights
> sum(designstatus$Weight)
[1] 15023.15
> sum(designstatus$final.wt)
[1] 7358.15
>
> # Write out Design Status File
> write.table(designstatus, file="designstatus.csv", sep=",", row.names=FALSE)
>

```